VISION-BASED NON-CONTACT DONKEY FACE LOCALIZATION AND INDIVIDUAL IDENTIFICATION USING IMPROVED YOLO11

/

基于改进 YOLO11 的非接触式驴脸定位与个体识别

Xinchao LI, Haojie ZHANG, Beihai ZHAO, Xin HE, Tingting ZHANG, Lijun CHENG*)
Faculty of Software Technologies of Shanxi Agricultural University, Shanxi / China;
Correspondent authors: Lijun CHENG: Tel: 13835441585; E-mail: cljzyb@sxau.edu.cn;
DOI: https://doi.org/10.35633/inmateh-78-119

Keywords: *donkey face recognition; individual identification; non-contact identity management; YOLO11; precision livestock farming*

ABSTRACT

Accurate individual identification is essential for precision donkey farming, but conventional methods based on manual records or physical tags are labor-intensive, inefficient, and vulnerable to tag loss or damage. To address these limitations, this study proposes MFW-YOLO11, a non-contact donkey face localization and individual identification model based on an improved YOLO11 framework. A total of 6,531 valid donkey face images were collected from a real farm environment and used to construct an individual identity dataset under diverse conditions, including different illumination, poses, occlusions, and backgrounds. In the proposed network, MANet is introduced into the backbone and head to strengthen fine-grained identity-related features, such as the eyes, nasal bridge, muzzle region, facial contour, and coat texture. A MANet-FasterCGLU composite module is further designed to adaptively filter effective facial responses and suppress interference from railings, donkey bodies, troughs, and complex backgrounds. In addition, a weighted feature union module is embedded in the neck to enhance the adaptive fusion of shallow texture details and deep semantic information. Experimental results show that MFW-YOLO11 achieves a precision of 90.7%, recall of 79.8%, mAP50 of 88.0%, mAP50–95 of 74.4%, FPS of 68.93, and GFLOPs of 6.3. Compared with the original YOLO11, the proposed model improves precision, recall, mAP50, and mAP50–95 by 6.5, 8.8, 6.2, and 6.4 percentage points, respectively, while maintaining real-time inference performance. These results indicate that MFW-YOLO11 provides an effective and practical solution for non-contact donkey individual identification in precision livestock management.

摘要

针对传统驴只个体识别方法依赖人工记录和物理标签、识别效率低且易受标签脱落影响等问题，提出一种基于改进 YOLO11 的非接触式驴脸定位与个体识别模型 MFW-YOLO11。研究在真实驴场环境下采集并筛选 6,531 张有效驴脸图像，构建覆盖不同光照、姿态、遮挡和背景条件的个体身份数据集。模型在 Backbone 和 Head 中引入 MANet 增强眼部、鼻梁、口鼻区域、脸部轮廓和毛色纹理等细粒度身份特征表达；构建 MANet-FasterCGLU 复合模块以自适应筛选有效驴脸响应，并抑制栏杆、驴身、饲槽和复杂背景干扰；在 Neck 中嵌入 WFU 模块以增强浅层纹理细节和深层语义信息的加权融合。实验结果表明，MFW-YOLO11 的 Precision、Recall、mAP50、mAP50-95、FPS 和 GFLOPs 分别达到 90.7%、79.8%、88.0%、74.4%、68.93 和 6.3，较原始 YOLO11 分别提高 6.5、8.8、6.2 和 6.4 个百分点。结果表明，该模型可为精准畜牧管理中的驴只非接触式个体识别提供技术支撑。

INTRODUCTION

As one of the world's major agricultural countries, China is rapidly developing large-scale and intelligent livestock production (*Li et al., 2021*). Accurate individual identification is therefore essential for improving production efficiency, disease prevention, and farm management. Donkeys have important draught, medicinal, meat, and milk values (*Zhu et al., 2025*), but their management still relies mainly on manual records, ear tags, collars, and other conventional methods. These methods are labor-intensive, susceptible to tag loss or damage, and limited in continuous tracking and automatic data accumulation (*Balieva et al., 2026*). With the development of precision livestock farming (PLF), reliable, efficient, and low-disturbance identification technologies are increasingly required for refined animal management (*Li et al., 2020; Papakonstantinou et al., 2024*).

RFID is commonly used through wearable devices such as ear tags and collars, and can record individual animal information under relatively controlled conditions (*Kang and Oh, 2025*). However, in real farms, RFID performance may be affected by electromagnetic interference, metal facilities, barn layout, reader deployment, and equipment maintenance, resulting in unstable signal transmission or recognition failure (*Long et al., 2025*). For donkeys with large movement ranges and variable behaviors, physical tags may also detach, be damaged, or cause discomfort, reducing the reliability of long-term management (*Mora et al., 2024*). In contrast, computer vision provides a non-contact and low-disturbance alternative, although traditional vision methods and biometrics such as iris or muzzle-print recognition still face limitations in robustness, acquisition distance, device requirements, and animal welfare (*Fuentes et al., 2023; Meng et al., 2025*).

Deep learning has promoted the application of facial recognition in animal individual identification. For pig face recognition, *Wang and Liu (2022)* proposed a two-stage method based on triplet margin loss and achieved an accuracy of 94.04% on the JD pig face dataset, demonstrating the potential of deep learning for precision pig management.

Wang et al. (2024) developed an unsupervised model for cattle individual identification, in which feature similarity matching was used to alleviate the transition from closed-set recognition to open farming scenarios. *Han et al. (2025)* introduced tracking technology and data augmentation strategies to address pose variations in dairy cow face recognition, achieving high recognition accuracy for 17 dairy cows.

Billah et al. (2022) combined the YOLOv4 detection algorithm with a convolutional neural network classifier to achieve effective goat individual identification.

For donkey face recognition, *Pan et al. (2025)* proposed a method combining the convolutional block attention module (CBAM) with ResNet50. Compared with the original model, the improved CBAM-ResNet50 increased accuracy and recall by 0.86 and 0.39 percentage points, respectively, while reducing the number of parameters by 72.42%. These studies indicate that deep learning-based facial recognition has promising potential in animal individual identification.

In practical donkey farms, surveillance images are often affected by viewpoint deviation, scale variation, uneven illumination, partial occlusion, and complex backgrounds (*Shu et al., 2026*). Moreover, individual donkeys usually have similar facial structures, and identity differences mainly appear in fine-grained regions such as the eyes, nasal bridge, muzzle, facial contour, and coat texture (*Hou et al., 2025*). To address these challenges, this study proposes MFW-YOLO11, an improved YOLO11-based model that treats each donkey as an independent identity category and integrates face localization with individual recognition. MANet is introduced to enhance fine-grained feature representation, MANet-FasterCGLU is designed to select identity-related features and suppress background interference, and WFU is used to improve multi-scale feature fusion. Experiments show that MFW-YOLO11 improves precision, recall, mAP50, and mAP50–95 by 6.5, 8.8, 6.2, and 6.4 percentage points over YOLO11, providing a feasible solution for non-contact donkey identification in precision livestock management.

MATERIALS AND METHODS

DATA COLLECTION SITE

The data used in this study were collected from Shanxi Yunkun Agriculture and Animal Husbandry Co., Ltd., located in Shangzhuang Village, Xiaobai Township, Taigu District, Jinzhong City, Shanxi Province, China. This area is situated in central Shanxi and represents a typical farming and livestock production region in North China. Taigu District has a warm temperate continental climate, with an average annual temperature of approximately 10.7 ℃ and an annual precipitation of about 423 mm. The relatively dry climate provides suitable conditions for livestock rearing, forage storage, and barn management.

Image acquisition was conducted in a donkey farm under real production conditions. The data collection scenes covered barns, exercise areas, and natural feeding areas. Unlike laboratory-controlled environments, the real farming environment involved natural illumination changes, complex backgrounds, freely moving animals, and partial occlusions. These factors better reflect the challenges encountered in practical applications and help evaluate the generalization ability of the proposed model under real-world farming conditions. The geographical location of the data collection site is shown in Fig.1.

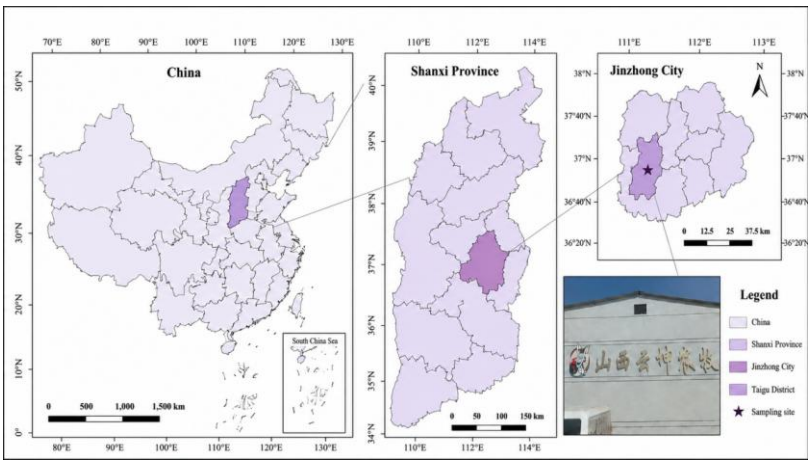
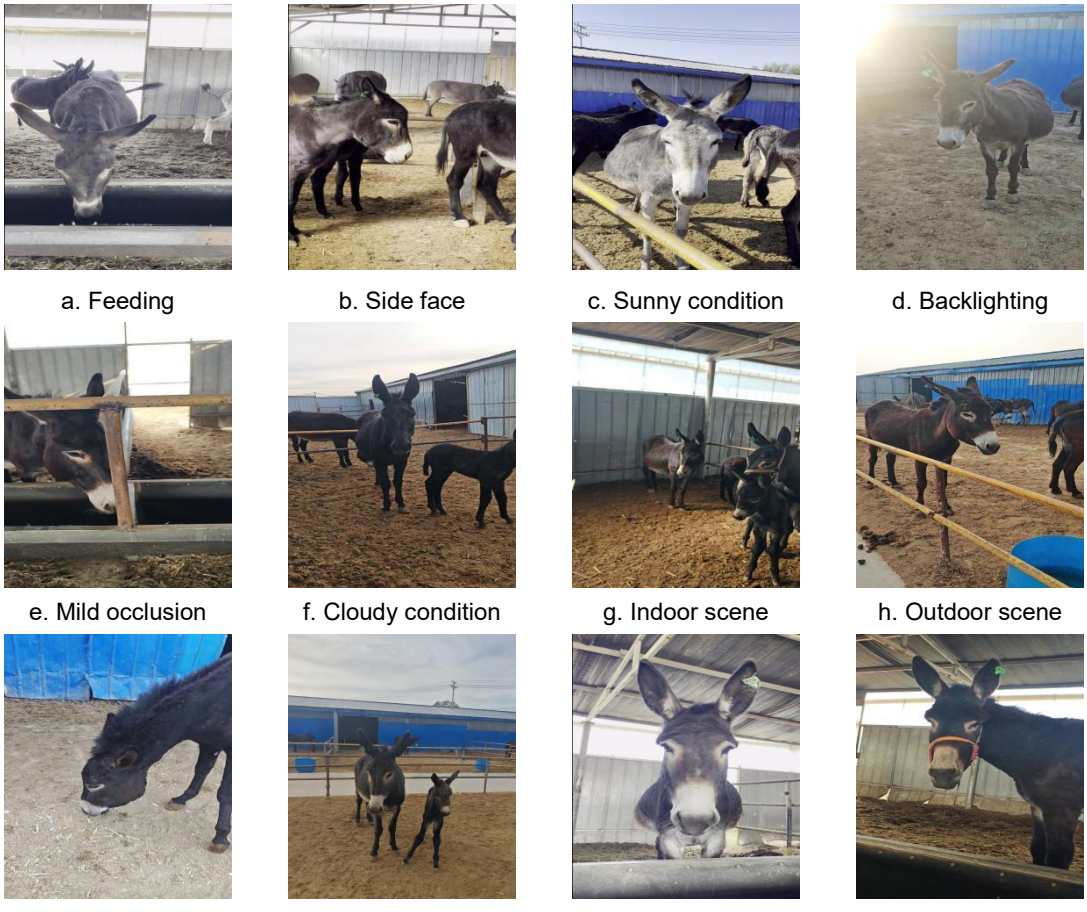


Fig. 1 - Geographical location of the data collection site

CONSTRUCTION OF THE DONKEY FACE INDIVIDUAL IDENTIFICATION DATASET

To address the limited data resources and insufficient consideration of complex natural farming environments in existing donkey face recognition studies, donkey face video data were collected under real farming conditions. Static image samples were then obtained through frame extraction. To avoid data leakage caused by consecutive video frames, the dataset was divided at the video-clip level, ensuring that consecutive frames extracted from the same video clip did not appear simultaneously in the training, validation, and test sets.

In the constructed dataset, each donkey was defined as an independent identity category. Therefore, the model output included both the donkey face bounding box and the corresponding individual identity label. The dataset covered various environmental conditions, including indoor and outdoor scenes, sunny and cloudy weather, front lighting, and backlighting. It also included multiple natural postures and behavioral states, such as frontal faces, side faces, head lowering, head turning, feeding, and mild occlusion, as shown in Fig.2.



a. Feeding b. Side face c. Sunny condition d. Backlighting

e. Mild occlusion f. Cloudy condition g. Indoor scene h. Outdoor scene

i. Head lowering j. Cloudy condition k. Frontal face l. Head turning

Fig. 2 - Examples of donkey face images under different farming conditions

All original images were manually screened for quality. Images were excluded if they were obviously blurred, if the donkey face was severely occluded by fences, other donkeys, or environmental objects, if the face region was too small, too dark, overexposed, or difficult to identify, or if more than 50% of the visible donkey face region was occluded. After frame extraction, quality screening, and data organization, a total of 6,531 valid images were finally obtained, as shown in Table 1.

Table 1

Dataset division for donkey face individual identification

Dataset	Number of images	Proportion	Purpose
Training set	5217	80%	Model training
Validation set	641	10%	Parameter optimization
Test set	673	10%	Model evaluation
Total	6531	100%	——

To further ensure the reliability and class balance of the donkey face individual identification dataset, the sample distribution of each donkey identity was statistically analyzed, as shown in Table 2. The dataset contained 27 individual donkeys, and each donkey was defined as an independent identity category. The number of images for each individual ranged from 213 to 262, with an average of approximately 241 images per donkey. This indicated that the dataset had a relatively balanced distribution among different identity categories, with no obvious class imbalance. In addition, during the division of the training, validation, and test sets, samples from each individual donkey were included in all subsets. This ensured that the model could learn the features of all identity categories during training and evaluate all individuals during validation and testing. Therefore, this data organization reduced the influence of sample distribution bias and provided a reliable data basis for subsequent model training, performance evaluation, and comparative experiments.

Table 2

Sample distribution statistics of the donkey face individual identification dataset

Statistical item	Result
Number of individual donkeys	27
Minimum number of images per donkey	213
Maximum number of images per donkey	262
Average number of images per donkey	241
Class imbalance	No
Samples of each individual included in training, validation, and test sets	Yes

IMPROVED YOLO11 MODEL FOR DONKEY FACE INDIVIDUAL IDENTIFICATION

To improve the accuracy and robustness of YOLO11 in donkey face individual identification under natural farming conditions, an improved model named MFW-YOLO11 was proposed using YOLO11 as the baseline. The proposed model was optimized from three aspects: fine-grained feature extraction, effective feature selection, and multi-scale feature fusion.

First, the MANet module was introduced into the deep feature extraction part of the backbone and the feature integration part of the head to replace part of the original C3k2 structures. MANet enhanced the extraction of fine-grained features, including donkey facial contours, coat texture, eyes, and muzzle regions, thereby improving the model's ability to represent identity-related differences among individual donkeys.

Second, the FasterCGLU structure was incorporated on the basis of MANet to construct a MANet-FasterCGLU composite module. Through a gated feature selection mechanism, this module strengthened the responses of effective donkey face features and suppressed irrelevant interference from railings, donkey bodies, troughs, and complex backgrounds. As a result, the model focused more effectively on key regions with identity-discriminative information.

Third, a weighted feature union (WFU) module was introduced into the neck to replace part of the original Upsample + Concat feature fusion strategy. The WFU module enhanced the fusion of shallow texture details and deep semantic information, thereby improving recognition stability under frontal faces, side faces, illumination variations, and complex background conditions.

Through the collaborative integration of MANet, MANet-FasterCGLU, and WFU, MFW-YOLO11 improved feature representation, interference suppression, and multi-scale feature fusion. Therefore, it was better suited for non-contact donkey face individual identification in real farming environments. The overall network structure is shown in Fig.3.

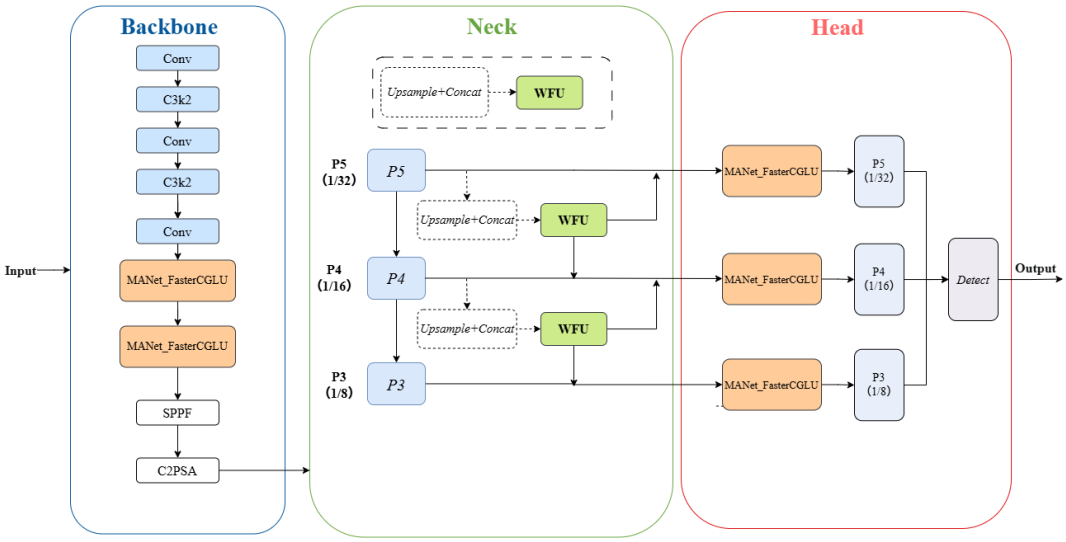


Fig. 3 - Network architecture of the proposed MFW-YOLO11 model

MANET

The MANet module was designed to enhance the network's ability to represent fine-grained facial features of donkeys. Because the facial appearance differences among individual donkeys were subtle, local features such as coat texture, eye shape, nasal ridge structure, and the mouth–nose region played a critical role in distinguishing individual identities. The original C3k2 module in YOLO11 exhibited limited capability to capture such local discriminative cues under complex backgrounds. Therefore, in this study, the MANet structure was incorporated into both the deep feature extraction layers and the detection head feature aggregation stage to improve the model's ability to extract multi-scale local features and regional structural information. Given the input feature map $X \in \mathbb{R}^{C\times H\times W}$, MANet extracted multi-scale features through multiple convolutional branches and an identity mapping branch, followed by channel-wise fusion. The feature aggregation process was formulated as shown in Eq.(1) :

$$F_{\mathrm{MAN}} = \mathrm{Conv}_{1\times1}\left(\mathrm{Concat}\left(\mathrm{Conv}_{3\times3}(X), \mathrm{Conv}_{5\times5}(X), X\right)\right) \quad (1)$$

Subsequently, a residual connection was applied to preserve the original input information and to reduce feature propagation loss in the deep network, as shown in Eq.(2):

$$Y_{\mathrm{MAN}} = F_{\mathrm{MAN}} + X \quad (2)$$

Here, $\mathrm{Conv}_{3\times3}$ and $\mathrm{Conv}_{5\times5}$ denoted convolutional feature extraction operations with different receptive fields, $\mathrm{Concat}(\cdot)$ represented channel-wise concatenation, $\mathrm{Conv}_{1\times1}(\cdot)$ was used to integrate multi-branch features and adjust the channel dimensions, F_{MAN} indicated the aggregated intermediate feature, and Y_{MAN} represented the output feature of the MANet module. By combining multi-branch feature aggregation with residual connections, MANet effectively captured local texture details and regional structural information, thereby enhancing the model's discriminative ability for visually similar donkey faces, as shown in Fig.4.

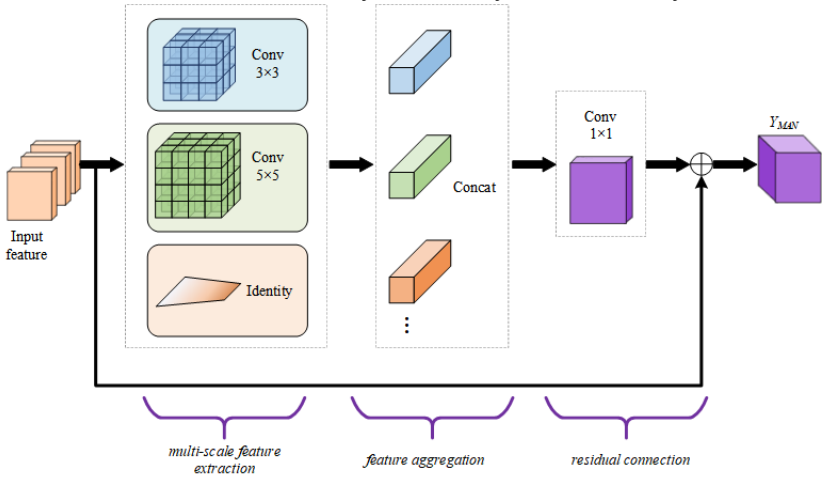


Fig. 4 - Architecture of the MANet Module

MANET-FASTERCGLU

FasterCGLU is a feature selection structure based on a gating mechanism. In this study, FasterCGLU was integrated with MANet to form a composite MANet–FasterCGLU module, which replaced certain C3k2 structures in the original YOLO11. This module first employed MANet to enhance the representation of fine-grained facial features of donkeys, including contours, eyes, nasal ridge, and mouth–nose regions. Subsequently, the gating mechanism of FasterCGLU adaptively selected the most informative features, thereby strengthening the key responses related to individual identity recognition while suppressing irrelevant information from backgrounds, fences, donkey bodies, and feeding troughs.

Let the enhanced features output by MANet be denoted as Z, which were obtained as described in the previous section, as shown in Eq.(3):

$$Z = Y_{\mathrm{MAN}} \tag{3}$$

Subsequently, FasterCGLU divided the input feature Z along the channel dimension into two sub-features, Z_1 and Z_2, as shown in Eq.(4):

$$Z_1, Z_2 = \mathrm{Split}(Z) \tag{4}$$

Here, Z_1 was used for efficient spatial feature extraction, while Z_2 was employed to generate the gating response. For the spatial feature branch, depthwise separable convolutions and 1×1 convolutions were applied to model local features, as shown in Eq.(5):

$$F_{\mathrm{a}} = \mathrm{Conv}_{1\times1}\left(\mathrm{DWConv}_{3\times3}\left(Z_1\right)\right) \tag{5}$$

For the gating branch, candidate features and gating weights were generated, and effective feature selection was performed via element-wise multiplication, as shown in Eq.(6):

$$F_g = W_v(Z_2) \odot \sigma\left(W_g(Z_2)\right) \tag{6}$$

Finally, the spatial feature branch F_a and the gating branch F_g were fused along the channel dimension, and a residual connection was applied to obtain the output of the MANet–FasterCGLU module, as shown in Eq.(7):

$$Y_{\mathrm{MFC}} = \mathrm{Conv}_{1\times1}\left(\mathrm{Concat}\left(F_{\mathrm{a}}, F_{\mathrm{g}}\right)\right) + Z \tag{7}$$

Here, $\mathrm{Split}(\cdot)$ represented the channel-splitting operation, $\mathrm{DWConv}_{3\times3}(\cdot)$ denoted the 3×3 depthwise separable convolution, $\mathrm{Conv}_{1\times1}(\cdot)$ indicated the 1×1 convolution, $W_{\mathrm{v}}(\cdot)$ and $W_{\mathrm{g}}(\cdot)$ corresponded to the candidate feature map and the gating weight map, respectively, $\sigma(\cdot)$ represented the Sigmoid activation function, $\odot$ denoted element-wise multiplication, and Y_{MFC} was the output feature of the MANet–FasterCGLU module. Through this architecture, the MANet–FasterCGLU module not only enhanced the fine-grained facial feature representation of donkeys but also adaptively emphasized discriminative identity features while suppressing interference from complex backgrounds, thereby improving the robustness of donkey face recognition in natural farming environments. The module architecture is illustrated in Fig.5.

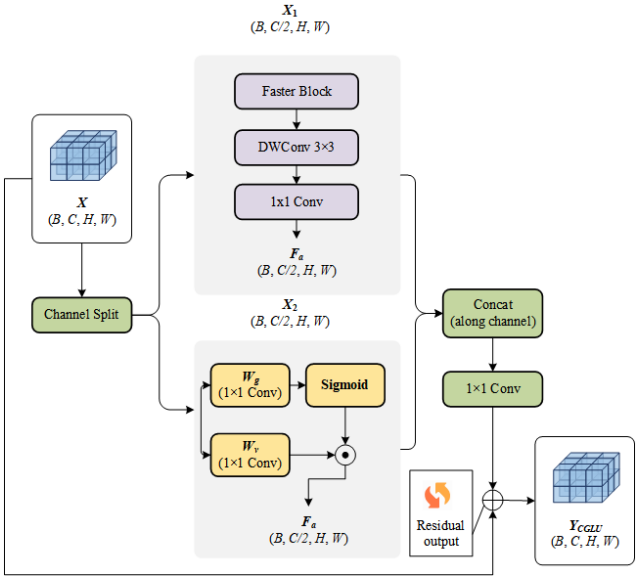


Fig. 5 - Architecture of the FasterCGLU Module

WFU

In the original YOLO11, multi-scale feature fusion in the Neck was typically performed using an upsampling and concatenation strategy. Although this approach was structurally simple, it lacked an explicit weighting mechanism between features at different levels, which often resulted in insufficient integration of shallow texture details and deep semantic information. To address this limitation, a Weighted Feature Upsampling (WFU) module was introduced along the top-down feature fusion path. This module enabled high-level semantic features and low-level detailed features to be adaptively fused according to the input content, thereby improving the model's robustness in recognizing donkey faces under varying poses, scales, and complex background conditions.

Let the high-level semantic features be denoted as F_h and the low-level detailed features as F_l. First, the high-level semantic features were upsampled to match the spatial dimensions of the low-level features, while the low-level detailed features were passed through a 1×1 convolution to align the channel dimensions, as shown in Eq.(8):

$$F_h^{'} = \mathrm{Up}\left(F_h\right), \quad F_l^{'} = \mathrm{Conv}_{1\times1}\left(F_l\right) \tag{8}$$

Subsequently, the aligned high-level and low-level features were concatenated and passed through a 1×1 convolution to generate fusion weights. To ensure the comparability of weights across different branches, the $\mathrm{Softmax}$ function was applied for normalization, as shown in Eq.(9):

$$[\alpha_h, \alpha_l] = \mathrm{Softmax}\left(\mathrm{Conv}_{1\times1}\left(\mathrm{Concat}\left(F_h^{'}, F_l^{'}\right)\right)\right) \tag{9}$$

Finally, the WFU module obtained the fused output features via weighted summation, as shown in Eq.(10):

$$Y_{\mathrm{WFU}} = \alpha_h \odot \mathrm{Up}(F_h) + \alpha_l \odot \mathrm{Conv}_{1\times1}(F_l) \tag{10}$$

Here, $\mathrm{Up}(\cdot)$ represented the upsampling operation, $\mathrm{Conv}_{1\times1}(\cdot)$ denoted the 1×1 convolution, $\mathrm{Concat}(\cdot)$ indicated channel-wise concatenation, α_h and α_l corresponded to the adaptive fusion weights for the high-level semantic and low-level detailed features, respectively, $\odot$ represented element-wise multiplication, and Y_{WFU} was the output feature of the WFU module. Through this weighted fusion mechanism, the WFU module simultaneously preserved the edges, textures, and local details of donkey faces while incorporating stronger semantic representations, thereby improving the robustness of donkey face recognition in natural farming environments. The module architecture is illustrated in Fig.6.

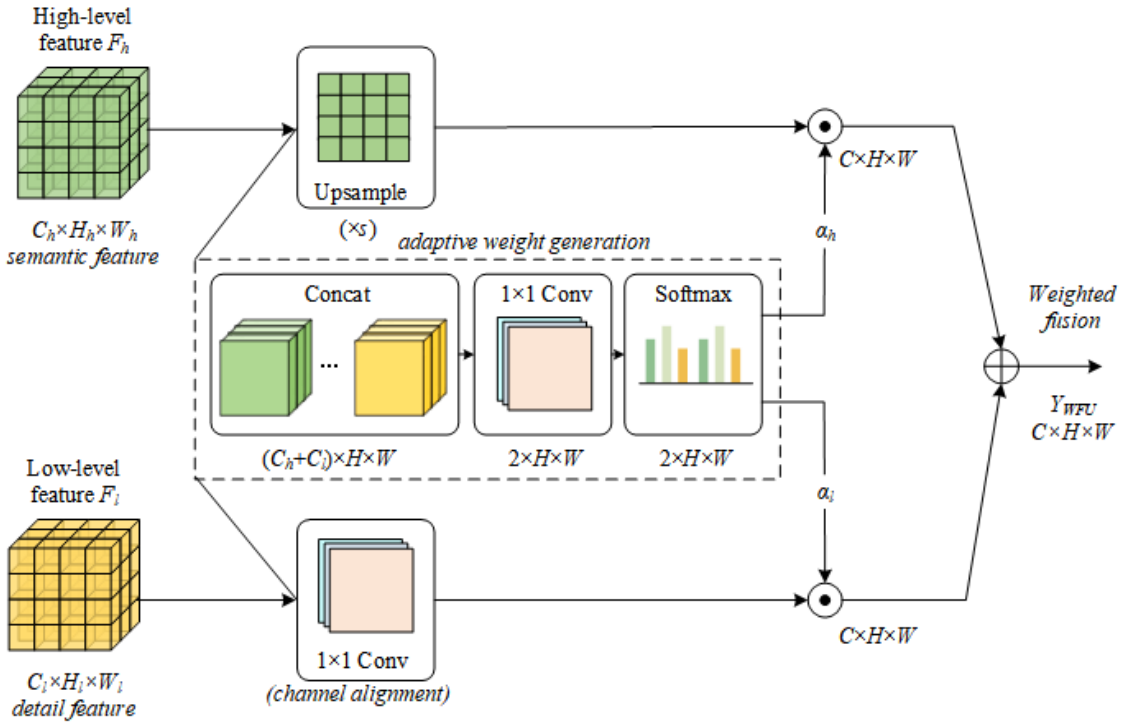


Fig. 6 - Architecture of the WFU Module

EVALUATION METRICS

In this study, precision, recall, mean average precision at an IoU threshold of 0.50 (mAP50), and mean average precision averaged over IoU thresholds ranging from 0.50 to 0.95 (mAP50–95) were adopted as the primary evaluation metrics. Specifically, mAP50 measured the average precision when the IoU threshold was set to 0.50, whereas mAP50–95 provided a more stringent and comprehensive assessment of detection performance across multiple IoU thresholds. The computation of each evaluation metric is presented in Eq. (11)-(14):

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$mAP_{50} = \frac{1}{C}\sum_{i=1}^{C} AP_i^{0.50} \tag{13}$$

$$mAP_{50:95} = \frac{1}{10}\sum_{t=0.50}^{0.95} mAP_t \tag{14}$$

EXPERIMENTAL SETUP AND PARAMETER SETTINGS

To ensure a fair comparison among different models, all experiments were conducted on the same workstation with identical hardware and software configurations. A detailed summary of the experimental environment is presented in Table 3.

Table 3

Experimental Environment Configuration

Component	Configuration
Operating System	Windows 11
CPU	Intel(R) Xeon(R) Platinum 8270 CPU @ 2.70GHz 2.70 GHz
GPU	NVIDIA GeForce RTX 4090 D
Programming Language	Python 3.10
Deep Learning Framework	PyTorch

The training parameters for each model are summarized in Table 4. To ensure comparability of the experimental results, input image size, number of training epochs, batch size, early stopping patience, and the number of data-loading threads were kept consistent across all models. Upon completion of training, the system automatically saved the model weight files, training logs, performance metric curves, and validation results, providing a basis for subsequent model performance analysis and comparative experiments.

Table 4

Parameter Settings

Parameter	Value
Input Image Size	640×640
Number of Training Epochs	250
Batch Size	32
Early Stopping Patience	30
Number of Data-Loading Threads	8

RESULTS

COMPARISON OF DETECTION PERFORMANCE

To verify the effectiveness of the proposed model in donkey face localization and individual identification, several mainstream detection models commonly used in animal face recognition studies were selected for comparison, including YOLOv5 (*Bergman et al., 2024*), YOLOv8 (*Ali and Muhammad, 2025*), YOLO12 (*Tian et al., 2025*), YOLOv13 (*Lei et al., 2025*), and RT-DETR (Real-Time Detection Transformer) (*Zhao et al., 2024*). The experimental results are presented in Table 5.

Table 5

Comparison of detection performance among different models

Model	P/%	R/%	mAP50/%	mAP50-95/%	FPS	GFLOPs
YOLOv5	85.1	61.1	72.7	59.5	73.96	7.7
YOLOv8	83.5	67.5	82.1	68.7	65.57	8.7
YOLO11	84.2	71	81.8	68	58.44	6.5
YOLO12	86.2	68.5	78.7	66.2	44.47	6.5
YOLOv13	82.8	63.6	74.8	62.8	31.53	6.4
RT-DETR	79.4	78.2	78.2	66.7	88.6	100.7
MFW-YOLO11	90.7	79.8	88	74.4	68.93	6.3

As shown in Table 3-1, the proposed MFW-YOLO11 achieved the best overall detection performance, with precision, recall, mAP50, and mAP50–95 of 90.7%, 79.8%, 88.0%, and 74.4%, respectively. Compared with YOLOv5, YOLOv8, YOLO11, YOLO12, YOLOv13, and RT-DETR, the model showed clear advantages in recognition accuracy and localization precision. Although RT-DETR obtained the highest FPS of 88.60, its precision and mAP values were lower than those of MFW-YOLO11, and its computational cost reached 100.7 GFLOPs. In contrast, MFW-YOLO11 maintained an FPS of 68.93 with only 6.3 GFLOPs, indicating a better balance among accuracy, speed, and computational cost for donkey identification in farming environments.

ABLATION STUDY

To verify the effectiveness of each improved module, ablation experiments were conducted using YOLO11 as the baseline model. Three single-module models were constructed by introducing MANet, MANet-FasterCGLU, and WFU separately. In addition, three dual-module fusion models were designed by combining MANet with MANet-FasterCGLU, MANet with WFU, and MANet-FasterCGLU with WFU. Finally, the complete model integrating all three modules was evaluated as MFW-YOLO11. The experimental results are shown in Table 6.

Table 6

Ablation experiment results of different module combinations

Model	Backbone		Neck	P/%	R / %	mAP50 / %	mAP50-95 / %	FPS	GFLOPs
	MANet	MANet_FasterCGLU	WFU						
1				84.2	71	81.8	68	58.44	6.5
2	√			87.4	77.5	87.1	72.3	54.7	8.4
3		√		91.4	70.5	81.4	67.9	66.36	8.9
4			√	89.6	77.3	86.4	72	62.01	8.1
5	√	√		90.3	73.3	84.6	70.9	59.38	7.2
6	√		√	87.9	76.7	86.5	72.4	56.66	10.2
7		√	√	89.9	77.8	86.1	71.8	53.59	8.9
8	√	√	√	90.7	79.8	88	74.4	68.93	6.3

As shown in Table 6, each improved module contributed differently to model performance. Compared with the original YOLO11, introducing MANet increased recall from 71.0% to 77.5% and improved mAP50 and mAP50–95 to 87.1% and 72.3%, indicating stronger extraction of facial contours, texture information, and local discriminative features. MANet-FasterCGLU alone achieved the highest precision among the single-module variants, indicating that the gating mechanism enhanced feature selectivity. However, its recall and mAP values did not improve significantly, suggesting that excessive feature filtering may suppress some difficult positive samples.

The dual-module combinations generally improved performance over the baseline, confirming the complementary effects of fine-grained feature extraction, effective feature selection, and multi-scale fusion. The complete MFW-YOLO11 model achieved the best overall results, with precision, recall, mAP50, mAP50–95, FPS, and GFLOPs of 90.7%, 79.8%, 88.0%, 74.4%, 68.93, and 6.3, respectively. Compared with YOLO11, these values increased by 6.5, 8.8, 6.2, and 6.4 percentage points for the four accuracy metrics, while FPS increased by 10.49 and GFLOPs decreased by 0.2 These results verify the effectiveness and synergy of the proposed modules.

VISUAL VALIDATION AND APPLICATION DEMONSTRATION OF THE SYSTEM

A visual application interface was developed in this study to further verify the practical applicability of the proposed donkey face individual identification model. The system was mainly used to demonstrate the recognition performance of the improved YOLO11 model on real donkey face images. As shown in Fig.7, the interface followed a complete workflow, including image upload, model recognition, result display, and heatmap analysis. Through this process, the system visually displayed detection bounding boxes, predicted identity categories, confidence scores, and heatmap responses of facial regions. These visualization results intuitively reflected the recognition effect of the proposed model and provided auxiliary support for non-contact donkey individual identification and precision management in farming environments.

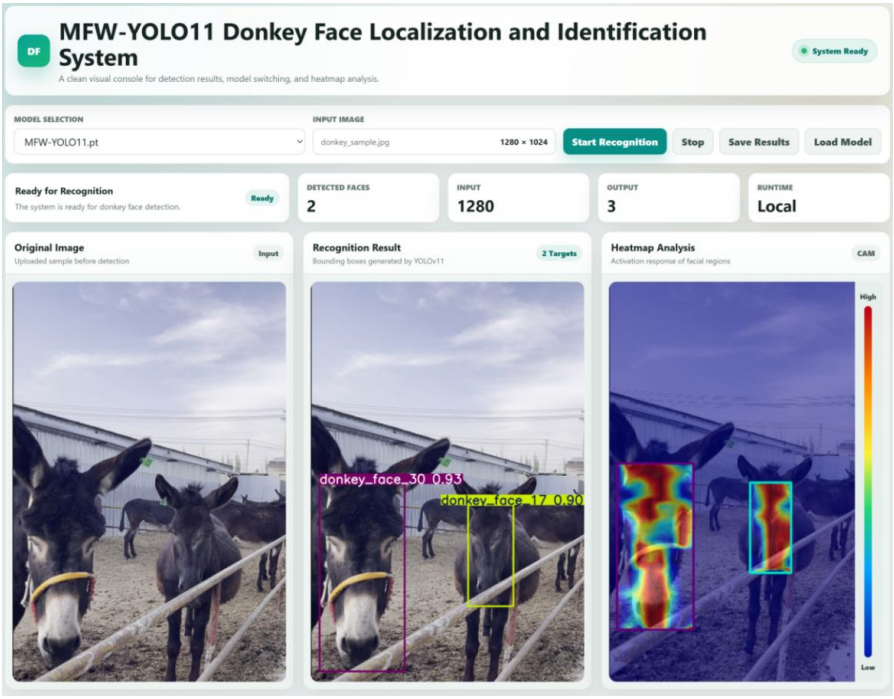


Fig. 7 - System Interface

CONCLUSIONS

This study constructed a donkey face individual identification dataset under real farming conditions to address the limitations of manual records and physical tags in continuous identity management. In the dataset, each donkey was defined as an independent identity category, enabling the integrated modeling of donkey face localization and individual identification. Based on YOLO11, an improved model named MFW-YOLO11 was proposed by introducing MANet, MANet-FasterCGLU, and WFU to enhance fine-grained identity feature extraction, effective feature selection, and multi-scale feature fusion. The proposed model made better use of discriminative facial information, including the eyes, nasal bridge, muzzle region, facial contour, and coat texture, while reducing interference from railings, donkey bodies, feeding troughs, and complex backgrounds.

Experimental results showed that MFW-YOLO11 achieved a precision of 90.7%, recall of 79.8%, mAP50 of 88.0%, mAP50–95 of 74.4%, FPS of 68.93, and GFLOPs of 6.3. Compared with the original YOLO11 and other representative detection models, the proposed model achieved a better balance among recognition accuracy, localization precision, inference speed, and computational cost, demonstrating its potential for non-contact donkey individual identification and precision livestock management. However, the current model is a closed-set identification method and can only identify donkey individuals included in the training dataset. Future work will expand the dataset across breeds, ages, seasons, and farming scenarios, and further investigate open-set recognition, multi-object tracking, identity database construction, and long-term cross-season identification for continuous video-based management.

REFERENCES

[1] Ali U., Muhammad W., (2025), Cow face detection for precision livestock management using YOLOv8, *International Journal of Innovations in Science & Technology*, vol.7, no.5, pp.128-132.

[2] Balieva G., Tanchev D., Lazarova I., Rankova R., (2026), Animal Identification in Precise Livestock Farming: A Systematic Review of Current Practices and Perspectives, *Kafkas Universitesi Veteriner Fakultesi Dergisi*, vol.32, no.2, pp.165-172. DOI: 10.9775/kvfd.2025.35661.

[3] Bergman N., Yitzhaky Y., Halachmi I., (2024), Biometric identification of dairy cows via real-time facial recognition, *Animal*, vol.18, no.3, article 101079. DOI: 10.1016/j.animal.2024.101079.
[4] Billah M., Wang X., Yu J., Jiang Y., (2022), Real-time goat face recognition using convolutional neural network, *Computers and Electronics in Agriculture*, vol.194, article 106730. DOI: 10.1016/j.compag.2022.106730.
[5] Fuentes A., Han S., Nasir M.F., Park J., Yoon S., Park D.S., (2023), Multiview Monitoring of Individual Cattle Behavior Based on Action Recognition in Closed Barns Using Deep Learning, *Animals*, vol.13, no.12, article 2020. DOI: 10.3390/ani13122020.
[6] Han S., Fuentes A., Park J., Yoon S., Yang J., Jeong Y., Park D.S., (2025), Utilizing farm knowledge for indoor precision livestock farming: Time-domain adaptation of cattle face recognition, *Computers and Electronics in Agriculture*, vol.234, article 110301. DOI: 10.1016/j.compag.2025.110301.
[7] Hou X., Huang X., Huang F., Dou Z., Zheng H., Wang C., Feng T., Liu M., (2025), A dataset of cow face and keypoint detection, *China Scientific Data*, vol.10, no.1. DOI: 10.11922/11-6035.csd.2024.0129.zh.
[8] Kang M.H., Oh S.H., (2025), Research trends in livestock facial identification: a review, *Journal of Animal Science and Technology*, vol.67, no.1, pp.43-55. DOI: 10.5187/jast.2025.e4.
[9] Lei M., Li S., Wu Y., Hu H., Zhou Y., Zheng X., Ding G., Du S., Wu Z., Gao Y., (2025), YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception, *arXiv preprint arXiv:2506.17733*. DOI: 10.48550/arXiv.2506.17733.
[10] Li B., Wang Y., Zheng W., Wang C., (2021), Research progress on intelligent equipment and information technology for livestock and poultry breeding, *Journal of South China Agricultural University*, vol.42, no.6, pp.18-26. DOI: 10.7671/j.issn.1001-411X.202107050. (in Chinese).
[11] Li N., Ren Z., Li D., Zeng L., (2020), Review: Automated techniques for monitoring the behaviour and welfare of broilers and laying hens: towards the goal of precision livestock farming, *Animal*, vol.14, no.3, pp.617-625. DOI: 10.1017/S1751731119002155.
[12] Long Y., Shi G., Cai J., Sun H., He L., Zhang X., (2025), Multipath-Rician channel interference evaluation model of RFID signal in pig breeding environment, *Transactions of the Chinese Society of Agricultural Engineering*, vol.41, no.20, pp.168-174. DOI: 10.11975/j.issn.1002-6819.202504175.
[13] Meng H., Zhang L., Yang F., Hai L., Wei Y., Zhu L., Zhang J., (2025), Livestock Biometrics Identification Using Computer Vision Approaches: A Review, *Agriculture*, vol.15, no.1, article 102. DOI: 10.3390/agriculture15010102.
[14] Mora M., Piles M., David I., Rosa G.J.M., (2024), Integrating computer vision algorithms and RFID system for identification and tracking of group-housed animals: an example with pigs, *Journal of Animal Science*, vol.102, article skae174. DOI: 10.1093/jas/skae174.
[15] Pan Y., Shen Y., Wang G., Zhang Y., Xu Z., Yu J., (2025), A donkey face recognition method combining attention module and ResNet50 model, *Heilongjiang Animal Science and Veterinary Medicine*, no.04, pp.125-130. DOI: 10.13881/j.cnki.hljxmsy.2024.08.0110. (in Chinese).
[16] Papakonstantinou G.I., Voulgarakis N., Terzidou G., Fotos L., Giamouri E., Papatsiros V.G., (2024), Precision Livestock Farming Technology: Applications and Challenges of Animal Welfare and Climate Change, *Agriculture*, vol.14, no.4, article 620. DOI: 10.3390/agriculture14040620.
[17] Shu H., Jin Z.M., Guo G., (2026), Deep learning-based Holstein face recognition in real-world farming conditions, *Smart Agricultural Technology*, vol.13, article 101690. DOI: 10.1016/j.atech.2025.101690.
[18] Tian Y., Ye Q., Doermann D., (2025), YOLOv12: Attention-Centric Real-Time Object Detectors, *arXiv preprint arXiv:2502.12524*. DOI: 10.48550/arXiv.2502.12524.
[19] Wang Y., Xu X., Zhang S., Wen Y., Pu L., Zhao Y., Song H., (2024), Adaptive group sample with central momentum contrast loss for unsupervised individual identification of cows in changeable conditions, *Applied Soft Computing*, vol.167, article 112340. DOI: 10.1016/j.asoc.2024.112340.
[20] Wang Z., Liu T., (2022), Two-stage method based on triplet margin loss for pig face recognition, *Computers and Electronics in Agriculture*, vol.194, article 106737. DOI: 10.1016/j.compag.2022.106737.
[21] Zhao Y., Lv W., Xu S., Wei J., Wang G., Dang Q., Liu Y., Chen J., (2024), DETRs Beat YOLOs on Real-time Object Detection, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.16965-16974. DOI: 10.1109/CVPR52733.2024.01605.
[22] Zhu Q., Khan M.Z., Peng Y., Wang C., (2025), A Comparative Review of Donkey Genetic Resources, Production Traits, and Industrial Utilization: Perspectives from China and Globally, *Animals*, vol.15, no.23, article 3372. DOI: 10.3390/ani15233372.